



UPPSALA  
UNIVERSITET

# Machine learning for covariate selection and prediction of pharmacokinetic data illustrated with rifampicin

Lina Keutzer, Huifang You, Ulrika SH Simonsson

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

Correspondance: [lina.keutzer@farmbio.uu.se](mailto:lina.keutzer@farmbio.uu.se)

## Introduction

In PKPD analysis, PK is used as input to drive the exposure-response relationship. This requires a population PK model, which can be time- and labor-intensive to develop. Machine learning (ML) stands out with its high computational efficiency and is being increasingly used in drug development. ML could thus support PMX with faster PK predictions used as input for a PKPD model, provided acceptable performance.

## Objective

In this work we explore the ability of different ML algorithms to 1) select predictors (= features) correctly and to 2) accurately and precisely predict plasma concentration over time using rifampicin as an example.

## Methods

Rifampicin plasma concentrations over time for 83 virtual patients were simulated from a previously published population PK model (1) with sample size and covariate distribution in accordance with the original study design (2). True predictors included in the simulations were time after dose (TAD), sampling week, dose and fat-free mass (FFM). The simulated dataset consisting of 1826 observations was split randomly into a training (80%) and a test dataset (20%) for model evaluation. Different ML algorithms were trained for selection of predictors (= features) and to predict rifampicin PK using 5-fold cross-validation, including Random Forest (RF), GBM, XGBoost and an ensemble model. Simulations were performed in NONMEM 7.3.0 (3) and ML model building in R version 3.6.3 (4).

## Results

All algorithms correctly selected TAD and dose as the most important predictors (Fig. 2). For PK prediction, XGBoost performed best with the highest  $R^2$  (0.70) and lowest imprecision (rRMSE: 63.7%). The ensemble model combining all three algorithms, as well as GBM also performed well with both having a  $R^2$  of 0.69 and a rRMSE of 65.2%. RF had the lowest  $R^2$  (0.66) and the highest imprecision (rRMSE: 68.7%). Graphical evaluation revealed a good predictive performance across all tested algorithms (Fig. 3). Runtimes were between 5.2 and 619.1 seconds depending on the ML algorithm.

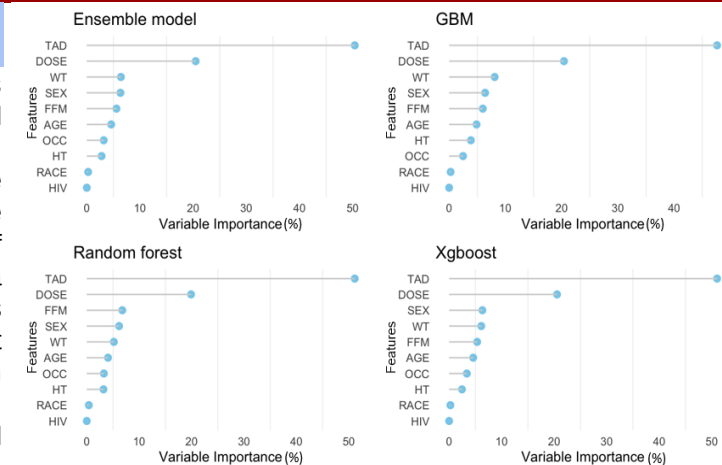


**Fig 2.** Feature importance scores shown for the ensemble model, GBM, Random forest and XGBoost. HIV: HIV-coinfection; OCC: Sampling week, RACE: race; HT: body height (cm); AGE: age (years); SEX: gender; WT: bodyweight (kg); FFM: fat-free mass (kg); DOSE: daily rifampicin dose (mg); TAD: time after dose (h).

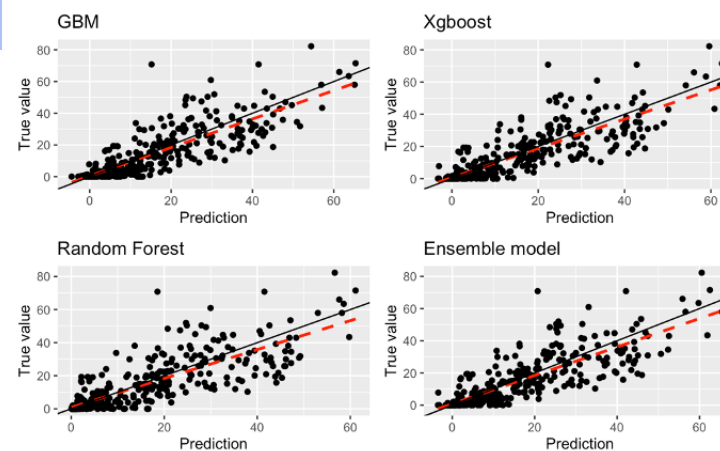


**Fig 3.** Predictions of rifampicin plasma concentrations from the different ML algorithms compared to the true observations. The black solid line is the line of identity, indicating 100% agreement between true and predicted values.

## Conclusions

This work shows that ML gives accurate PK predictions and provides a faster alternative to population PK as input to PKPD analysis when simulations are not needed. The increased efficiency of ML is not only due to the faster runtimes, but perhaps mostly the lack of need for structural model development. Bridging PMX and ML seems very promising considering that ML can add value to PMX workflows through increased efficiency.

## References

- [1] Svensson *et al.* A Population Pharmacokinetic Model Incorporating Saturable Pharmacokinetics and Autoinduction for High Rifampicin Doses. *Clin Pharmacol Ther* **103**, 674–683 (2018).
- [2] Boeree *et al.* A dose-ranging trial to optimize the dose of rifampin in the treatment of tuberculosis. *Am. J. Respir. Crit. Care Med.* **191**, 1058–1065 (2015).
- [3] Beal S, Sheiner L, Boeckmann A, Bauer R. NONMEM 7.3 Users Guides. Gaithersburg, MD.: ICON plc; 1989. Available from: <https://nonmem.iconplc.com/nonmem730/guides>.
- [4] R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2015).

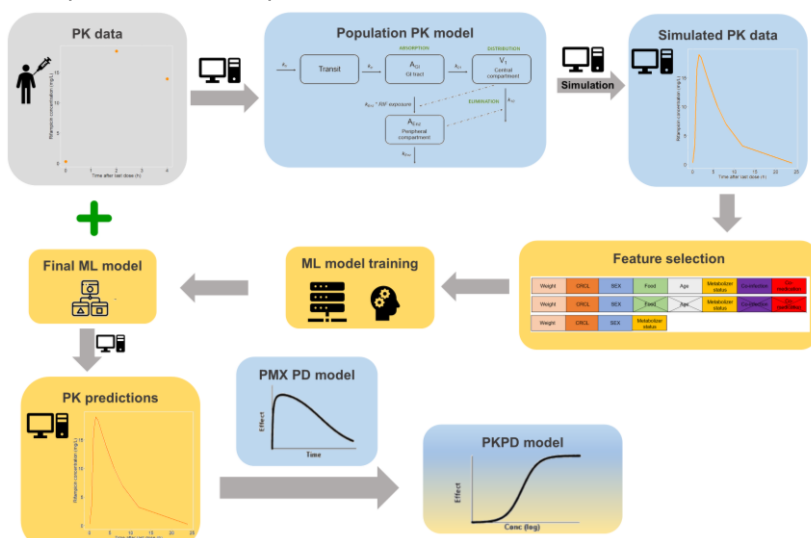


**Fig 1.** Workflow. Blue represents pharmacometrics and yellow machine learning.